Indraprastha Institute of Information Technology Delhi (IIITD)

Department of Computational Biotechnology

BIO213 – Introduction to Quantitative Biology

ASSIGNMENT-2

Name-PARAS DHIMAN || Roll no-2021482

Ques1:

a. A well commented python code has been provided

b. The Output is Given as :

```
These are the sequences printed 50 in each line:
1       SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDML
        -HHHHHHHHH-HSSSSSSSSSSSSSSHHHHHHSSSSSSSSHHHHHHHHHH
2       NPNYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPK
        H--HHHHHHHHHHSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS----HS
3       YKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGF
        SSSSSSSSSSSSSSSSSS--SSSSSSSSS---HHHHHH---------
```

Ques2:

Secondary structure prediction is an important aspect of protein structure determination. There are two categories of secondary structure prediction methods/tools, which are Chou and Fasman and GOR method (Category I) and DSSP, P-curve, and Stride (Category II). The major differences between these two categories are discussed below:

- Algorithmic approach:
  The Chou and Fasman and GOR method use an algorithmic approach to predict secondary structure. These methods rely on a set of empirically derived rules to predict the likelihood of each amino acid in a protein sequence to adopt an alpha helix, beta sheet, or random coil secondary structure. On the other hand, DSSP, P-curve, and Stride use a more sophisticated approach that involves assigning secondary structure types based on the hydrogen bond patterns observed between amino acid residues.

- Input data:
  The Chou and Fasman and GOR method require only the primary sequence of the protein as input. They are, therefore, faster and more convenient to use than the other methods in Category II. In contrast, DSSP, P-curve, and Stride require atomic coordinates of the protein

structure, which means that their use is limited to experimentally determined protein structures.

- Accuracy:
  The accuracy of secondary structure prediction varies between the two categories of methods. The Chou and Fasman and GOR method, while fast, are less accurate compared to DSSP, P-curve, and Stride. The latter methods can accurately assign secondary structures to individual amino acids in a protein and provide information on the overall conformation of the protein.

- Software availability:
  There are a large number of web-based and standalone software tools available for secondary structure prediction using Chou and Fasman and GOR method. In contrast, fewer tools are available for secondary structure prediction using DSSP, P-curve, and Stride. However, the latter methods are often integrated into larger software packages for protein structure determination and analysis.

Category I: Chou and Fasman and GOR method

The Chou and Fasman method and the GOR method are two examples of prediction methods that use statistical analysis of known protein structures to predict the secondary structure of a given protein sequence. These methods are based on the observation that certain amino acid residues have a higher propensity to form particular secondary structures, such as alpha-helices, beta-sheets, and turns. The Chou and Fasman method and the GOR method assign probabilities to each amino acid residue to determine the likelihood of it forming a particular secondary structure. They use a sliding window approach to identify regions of the protein that are likely to form specific secondary structures based on the amino acid sequence.

Category II: DSSP, P-curve, and Stride

The DSSP, P-curve, and Stride methods are examples of prediction methods that use a combination of empirical rules and knowledge-based algorithms to predict the secondary structure of a protein. These methods analyze the spatial arrangement of the protein's atoms to identify patterns that are characteristic of certain secondary structures. They also take into account factors such as hydrogen bonding, backbone torsion angles, and amino acid sequence patterns to refine their predictions.

One of the major differences between the two categories of secondary structure prediction methods/tools is that Category I methods rely more heavily on the amino acid sequence to predict the secondary structure, while Category II methods take into account the spatial arrangement of the protein's atoms. Category II methods are generally considered to be more accurate than Category I methods because they use more information, such as the actual 3D structure of known proteins, to make their predictions.

In conclusion, both categories of secondary structure prediction methods have their advantages and limitations. The choice of method largely depends on the availability of input data and the level of

accuracy required for the analysis. Category I methods are based on statistical analysis of known protein structures and assign probabilities to each amino acid residue to predict the secondary structure. Category II methods use a combination of empirical rules and knowledge-based algorithms to predict the secondary structure based on the spatial arrangement of the protein's atoms, as well as other factors such as hydrogen bonding, backbone torsion angles, and amino acid sequence patterns. Category II methods are generally considered to be more accurate and less prone to false positives than Category I methods.